Wide Area View

An interview with WAIS, Inc. CEO Brewster Kahle. By Jeff Ubois

Brewster Kahle is founder and CEO of WAIS, Inc., the San Francisco-based developer of Wide Area Information Server (WAIS) database search software. Founded in July 1992, WAIS employs more than 40 people. The company was recently acquired by America Online (AOL) for $15 million. Kahle, who in the mid-1980s was the architect of the CPU of the Thinking Machines Connection Machine Model 2, said his company's goal was to create tools that help people become publishers on the Internet.

While WAIS's primary products are free and commercial versions of its software, the company's publication services are growing rapidly and now account for about one-half of its business.

As part of AOL's flotilla of Internet software companies, WAIS is positioned between consumer online services and the traditional Internet culture from which it evolved—a potentially prime place to catalyze the emerging electronic publishing market.

The company's Web page is at **http://www.wais.com**, and the newsgroup **comp.infosystems.wais** offers helpful information for putting up WAIS servers.

IW: *What is WAIS, the software?*

KAHLE: WAIS is an Internet tool for searching for information. Right now the Internet is getting a lot of attention because it's not just a vaster wasteland; it offers the opportunity to participate. It's not just about getting information, but about being able to make your words known. So I think of the Internet as publishing, and the focus we have is on helping people say what they want and finding others that have similar interests.

We like to use this analogy: A book has three sections—the table of contents, the pages, and the index. So think of the Internet as a book: There is Gopher, which is the table of contents; there is the World-Wide Web, which is the hypertext pages; and there is WAIS, which is a directed search when you know what you want. WAIS is for the user who knows what he or she wants, not for the explorers or the tourists so much as for the people who want their answers. The index is what lets the reader take control.

IW: *How is WAIS different from traditional search engines?*

KAHLE: We provide for a user-friendly environment by accepting natural language questions, but we also provide for the trained searchers that know how to use Boolean logic and fielded searches. We see the amount of information on the Net growing phenomenally, so having assistance in finding things appropriate for you is crucial. We use a technique called relevance feedback, which is having the machine understand what you liked and didn't and using that data to find more documents. You say, "I like that one, find me more like that one." The machine looks at how you used various documents—what you read all the way through, what not, and what you cut and pasted and forwarded to friends.

IW: *Who's buying your products and services?*

KAHLE: Our main markets are government, which has a mandate to make information available for free; publishers, which have the traditional role of understanding how
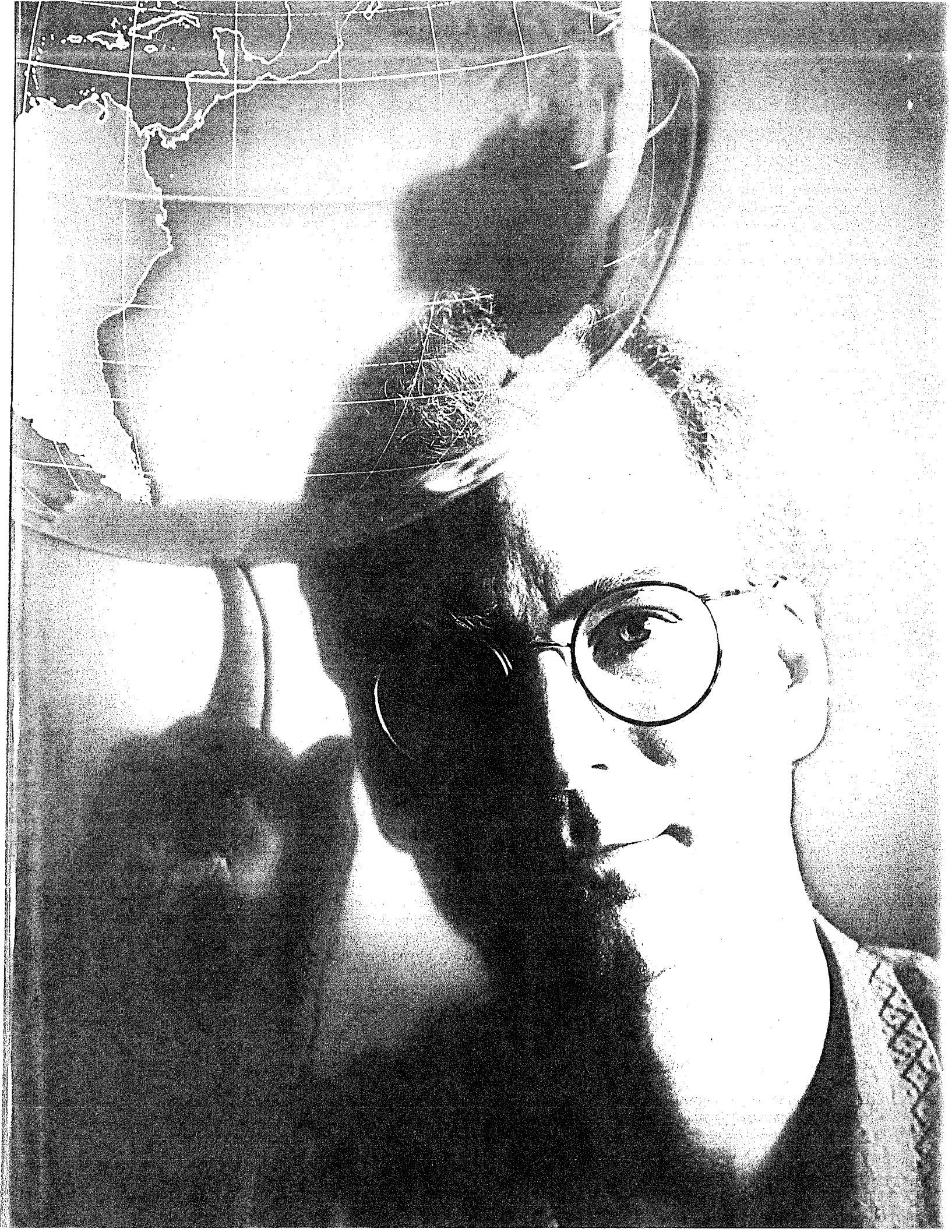
I keep all the mail I have sent or received. I index it every night and search back on that. It is the thing I use WAIS for most.

to distribute information for money; libraries; and distributed corporations—they are growing more and more global all the time and just staying in touch with themselves and finding resources in themselves becomes more and more difficult. Customers include the Library of Congress, Encyclopedia Britannica, Scholastic, Dow Jones, the Defense Technical Information Center, Perot Systems, and the intelligence community.

IW: *How many WAIS databases are out there now, and how many people are using them?*

KAHLE: We don't know. There are thousands of databases, and most users have used WAIS but probably don't know it.

IW: *Would you expand on the relevance feedback and natural language ideas?*

KAHLE: As databases are starting to be used by more and more people, many of them are not trained in Boolean logic, and when we look at the searches they do, they often use only one or two words. Trying to find the right document out of 100,000 documents based on one or two words is extremely difficult.

Anything we can do to help people tell us what it is they are looking for helps. From a list of documents, you can click what you like and it uses relevance feedback for you.

There are lots of other things that are being built around WAIS—not the core technology, but tools that let you hook multiple databases together. Companies like PLS (**http://www.pls. com**) are developing mechanisms for aggregating multiple databases, and Z39.50 [a library catalog access protocol] continues to spread.

All the companies are starting to see that the threat is not each other, but trying to make a system on the whole good enough that the Internet will hold enough value to keep proprietary systems at bay.

IW: *How do the search tools relate to the size of the database? If you've got 100 records, you don't care as much about the search tool as when you have 30 gigabytes.*

KAHLE: Absolutely. But it is key to have one interface that can address large numbers of databases. And not just external databases: Your own personal e-mail and your corporate files as well as wide-area information should be accessible from one point-and-click interface.

Most people find their own information to be the most important, their group's information somewhat less important, and wide-area information even less. But the amount of information available goes the other way. So the tools have to get more and more sophisticated the further away you get from people. Finding what you want among terabytes of data requires serious tools.

IW: *How can people use WAIS for personal data?*

KAHLE: I keep all the mail I have ever sent or received. It's about 500 megabytes. I index it every night to keep it up to date, and I use it as my memory. When I am trying to remember a name, I will send myself e-mail and I can search back on that. It is the thing I use WAIS for most. It is common to save all the messages you have ever sent and ever received, and you will start to save all the documents you have ever read from different sources.

IW: *What do you see as the most interesting emerging publishing technologies? Agents?*

KAHLE: A lot of our sophisticated customers are asking what comes next, what is after the Web. Everyone is getting increasing usage, but they are asking their professional friends and finding that people are not going back to their sites over and over again. So how do you raise the value so that people return to your site?

We are working with different technologies to deliver these types of capabilities. The simplest is personal pages, where you do a search and save and the next time you come online it has the page. That is personalization, but it doesn't have the other two aspects of an agent—aggregation and packaged delivery.

Aggregation is in its infancy right now, and packaged delivery is also still pretty wimpy. We've got e-mail delivery,

like Newshound [on AOL]; we have fax delivery, like First! (a service of Individual Inc; see below), and we are starting to get more custom applications like Ensemble's personal newspaper or Pointcast, where it is a little more layout oriented, oriented towards online browsing. We are working with Ensemble (**http://www.ensemble.com**) and building two different publishing systems using that technology.

IW: *What is Ensemble?*

KAHLE: Ensemble is a small development company that has been working on personal digital newspapers. They were distributing a *Wall Street Journal* for free on the Internet until recently. Dow turned them off because they were getting too popular. The technology is a digital newspaper that might have the same content for everybody—whether a *Wall Street Journal* or a *New York Times*—or it can be personalized. I personalize it to watch for certain companies, categories, and industries, and I receive it via e-mail.

IW: *I've been disappointed with clipping services. They give you too much garbage, or miss important stories, or filter out randomness. Is that getting better?*

KAHLE: No, not yet. But profiling what a user wants is difficult. If you ask them what they are interested in, often it is not an accurate assessment of what it is they will actually read if you hand it to them. We have not gotten much beyond having you fill out a form and sending documents based on those words, phrases, categories, and keywords. Right now there are two systems to get beyond that: One is human intervention, which is expensive but good, like Individual Inc. (**http://www.individual. com**), and the other is deduction based on people's interests.

The best work is taking place at MIT. The Homer/Ringo music rating system (**http://jeeves.media.mit.edu/ringo**) and the video ranking system at Bellcore (**videos@bellcore.com**) are good examples. Bellcore's system asks you to rank videos you like and it will recommend other videos based on the rankings of other people who are similar to you. These systems

Whether pricing will be à la carte, fixed fee, or per document is anyone's guess. Smart people are playing it a few different ways.

have extremely high predictive value.

So, aggregation is getting fixed by open protocols being adopted by a growing number of players, and personalization is getting better. Currently these are the bottlenecks.

For packaged delivery, we are getting some imaginative work in Silicon Valley to produce packaged applications for vertical markets. If you are a stock broker, for example, you would want graphs and reports and numbers; and if you are a CEO, you would be looking for press releases and articles about competitors, and you may want it to be packaged in a different way.

IW: *For a while it seemed like Gopher, WAIS, and the Web were all being discussed as tools everyone would use, but as forms-based Web pages become more common, do you see WAIS retreating into the woodwork?*

KAHLE: If we are successful, no one will know they are using WAIS. We are part of the plumbing, and work back upstream as a server technology. We try to be agnostic about information delivery systems. The Web is certainly the dominant delivery system now, but we see possibilities in the 3D environments, the Microsoft Network, personal digital newspapers—all sorts of different delivery technologies. All we are interested in is being on the back end, and the only time you notice the plumbing is when it backs up. So our goal is to be out of sight and working.

IW: *Are there any innovations in the copyright field you are watching, or that you think are interesting and relevant to publishers?*

KAHLE: It's funny, we don't end up in a lot of the same conversations we did ten years ago. Ten years ago publishers were very scared of having an article taken and sent around to all your pals. This doesn't seem to be high on their worry lists now.

IW: *I'm surprised to hear you say that. ClariNet lost Dave Barry because people were forwarding the articles around.*

KAHLE: Dow and CMP and others would care if the information were hoarded and resold. Dave Barry may have been an issue because there were Dave Barry lists where the articles were delivered to thousands and thousands of people. If it's just somebody taking a cool article from *Internet World* and sending it to their buddy, this doesn't seem to make publishers see red. If people republish it, then, heck, come down on them.

It is a matter of scale. The publishers we work with most actively are daily, weekly, and monthly publishers, and their value is in timeliness, completeness, and quality. Timeliness and completeness aren't served by someone sending around an article. They actually are interested in accumulating more users and will offer articles for free in the hope that people will subscribe to the system.

IW: *How do you see chargeback mechanisms evolving and how central is that to what you want to do?*

KAHLE: There are several competing systems—CyberCash, DigiCash, and others—and we see those as extremely exciting. We haven't worked with them yet, but we look forward to doing so in the future. Whether it will be *à la carte* pricing, fixed fee, or per document is anyone's guess, and the smart people are playing it a few different ways.

Right now, the simplest fee structure is subscription-based. So you can pay for a month of access, maybe a site license, and it's all you can eat, like ClariNet. It is an effective pricing model because people don't quite know what the value of the information is; they don't know how much they'd use it. What we are oriented towards is not even setting up shops so we make money. We are setting up shops so that others can make money.

Having people make money by publishing is crucial to the success of the network publishing system. But there is a culture clash here: While most people would be perfectly happy to give their words away for free, there are some words that are worth paying for, and we could all access those if a good chargeback mechanism was in place.

IW: *How do you compare WAIS to other forms of electronic distribution, such as*

CD-ROM and the older, non-consumer online services?

KAHLE: WAIS is very much on the publishing model. WAIS publishers control the distribution of their work so that they can have people subscribe, or have 30 days for free, or make it so users can see headlines but not the documents. They can arrange their business model in many different ways.

It is advantageous over CD-ROM because it is very easy to make updates, and to make much larger collections of info available. Unlike CD-ROM, where you are creating all content and giving it to people and hopefully they don't abuse it, you have control. It is easier to distribute to those people because it is using a shared network backbone.

Disadvantages? We don't have the bandwidth that CD-ROM has between the disk and the screen. That is a limitation compared to CD-ROM.

In comparison with Dialog or Mead Data, what people are looking for is to use the power of their desktop machine instead of just as a dumb terminal dial up, so there are those services now available on the Internet.

Another difference is cost. Where mainframes cost millions, putting out a CD can cost $100,000. With networks, to make yourself a network publisher, costs between $10,000 and S50,000. For that you can reach a worldwide audience serving thousands of users a day. This change enables many more people to become publishers. That is where we see the exciting aspect of this.

IW: *WAIS the company and WAIS the search tool have been strongly identified with the Internet, but now you've been acquired by America Online. From your perspective, how is the relationship between the online services and the Internet changing, and how will it affect publishing on the Net?*

KAHLE: WAIS technology is for publishing on the Internet, which has grown in importance in the last few years. AOL's interest in WAIS is in pushing "open" technologies to hedge against upcoming proprietary systems. "Publishing" has

grown more complicated as more data in different formats are processed and user expectations rise. We help Net publishers with their problems—from data handling to billing to advertising. AOL helps us with this through funding, technology, and access to their customer base.

The challenge the Internet has raised is one of open vs. closed, participatory vs. proprietary. The Internet is a celebration of the open, and we are getting our chance. There is a huge investment in time, infrastructure, and public training going into the Internet. But if we don't find a way to make this investment pay back, the experiment will die.

Maybe surprisingly, that is why we sold WAIS Inc. to America Online. Once a closed system, America Online has decided the best way to win as a minority player is to raise the quality of the Internet as a defense against upcoming proprietary online systems. AOL is large enough to have an impact, but small enough to think it cannot monopolize.

At that point, WAIS Inc.'s mission to "help publishers make money by publishing on the Internet" became strategically important to AOL. To WAIS, an acquisition meant enough leverage to deliver on a larger scale.

If the Internet community loses, then proprietary systems will replace the open ones: proprietary payment systems, proprietary page layouts, proprietary three-dimensional chat protocols, proprietary "extensions." WAIS Inc. is here to enable publishers to more than make back their investments. An open Internet is the best way to achieve this.

IW: *Can you talk about the AOL deal?*

KAHLE: We are a wholly owned subsidiary, a separate company, with the same charter for producing products and production services. What we get out of it is two-fold: one is the resources to grow well. Even though we were tripling each year profitably, there are corners you cut when you are bootstrapped, and this allows us to do a better job with our mission. The other aspect of the acquisition is we are tying into a larger organization that has a large user base.

We are still a company that serves publishers so their information can be accessed from any network, not just AOL; but we can leverage AOL to prove some of the business models that eventually this whole industry will adopt.
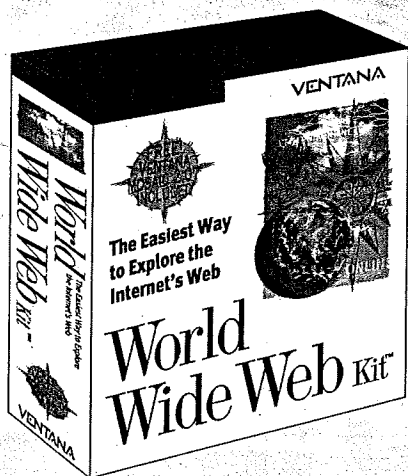
IW: *There is a lot of animosity and jeering towards AOL by users on the Internet. Why is that?*

KAHLE: You can paint AOL in a different light than the Internet community sees them now. There is the idea that AOL is just sponging off the Internet, and what we see is a change. AOL's approach was to gateway to Internet functions, but by buying and funding companies such as WAIS and others with an open charter, this is creating a different AOL. It is not done as altruism, but strategically for all the right reasons. The fight the Internet community has had to keep open standards is being embraced by AOL.

AOL is bringing on board a consumer group that I see as healthy for the Internet, and making dial-up access to the Internet understandable to larger communities of people. If the Internet wins, AOL wins.

IW: *AOL as a white knight for the Internet?*

KAHLE: They don't see themselves as big enough to be a monopoly, and what you do when you aren't big enough to be a monopoly is to work on the part of open systems. That is how Sun ate into DEC's market.

There are only a few players that are big enough to think they can dominate this area. Microsoft, maybe, and I would imagine the phone and cable companies will try. So what we are trying to do is to get an open infrastructure to be good enough at searching and finding things to promote the open technologies.

IW: *So what is AOL's strategy in buying you?*

KAHLE: Their purpose in buying WAIS, GNN, Medior, and Webcrawler is a strategy to raise the quality of the open Internet and a strategy to compete with MSN [Microsoft Network].

IW: *What are you going to do about MSN?*

KAHLE: We see MSN as another distribution channel, and to the extent we can help our publishers publish on MSN as well as the open Internet, that will be our strategy. The best argument for openness is quality. And we are here to make the quality piece go forward; and AOL sees that as a strategic requirement to battle Microsoft Network.

We don't see the world condensing into AOL and MSN information providers and readers. We see this as staying open so there are information providers and there are readers. And what our publishers want is one-stop shopping to make their information available to as many customer sets as possible.

IW: *Steve Case has made some very strongly worded statements against Microsoft in which he compared the operating system to the dial tone of a phone and a common carrier, and claimed that bundling MSN with Window 95 was denying consumers choices.*

KAHLE: I actually think he is right. There are some common-carriage issues there.

People very often use the software that is bundled, and as computers mature that seems to be happening at an even faster pace, so that either you are bundled or have a niche market. Microsoft has a strong position in the bundled market.

IW: *What about working with other AOL acquisitions? Will GNN be publishing with WAIS? Will BookLink build in special WAIS hooks?*

KAHLE: We'll be working more with GNN. The acquisition is still very young, and we sister companies are in the process of understanding each other to figure out how we can pull together new services.

IW: *Where do you think all that puts your company in one to three years?*

KAHLE: There is always room for high-end systems for the best publishing tools and services. We are finding that our publishing partners are realizing how difficult it is to do a world-class product. It used to be enough to throw your documents into a Gopher and you were done. Now you

need forms-based input, and e-mail feedback, and sophisticated services.

IW: *When you look five or 10 years out, who do you see publishing on the Net and who do you see generating data—individuals, organizations, automated systems like weather satellites? Which of those are your customers and what are you doing for them on the publishing side?*

KAHLE: We are serving those who are trying to repurpose their publications. Like McLuhan said, the new medium contains the old medium. It's not because there aren't clever people out there who want to use new media in new and different ways; it's just that someone has already paid for content to be built. So most of our customers are currently in that camp.

There are growing numbers of companies, though, that are specifically targeted for the Internet, which is not really cross-purposing. Time Inc. has 100 journalists working on Pathfinder, which is a large operation, and it indicates the cost of generating new content. Also, the Ringo system is being commercialized, so they are in startup mode. That is completely new and different and couldn't have happened before.

There is a program AOL has called the Greenhouse, to help info-entrepreneurs create their dream by seeding them with small amounts of money to help them get out there.

So we are seeing a new wash of content coming onto the Net, and it is often not straight text; it is weaving together multiple sources, like Yahoo or Webcrawler, that are serving a very useful purpose and a new medium that does not have final print or publication.

IW: *Any other thoughts on your future economic opportunities on the Net ?*

KAHLE: The people that are making a lot of money right now are the plumbers—people who are making bandwidth available, so cellular, cable, and phone companies all are going nuts in terms of increased demand. Next comes the information services groups, and that is what we are oriented towards. We'd like to get at Sega and Nintendo users and those set-top boxes that are starting to network. Hooking databases in the back end of such devices will bring a phenomenal opening up of applications. ∎